BEBR

FACULTY WORKING
PAPER NO. 1130

Nes[illegible] and Non-nested Procedures for Testing Line[illegible] [illegible]d Log-linear Regression Models

Anil [illegible]
Mich[illegible]

College of Commerce and Business Administration
Bureau of Economic and Business Research
University of Illinois, Urbana-Champaign

BEBR

FACULTY WORKING PAPER NO. 1130

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

March, 1985

Nested and Non-nested Procedures for Testing Linear and Log-linear Regression Models

Anil K. Bera, Assistant Professor
Department of Economics

Michael McAleer
Australian National University

Nested and Non-nested Procedures for Testing
Linear and Log-linear Regression Models*

Anil K. Bera
Department of Economics
University of Illinois

Michael McAleer
Department of Statistics
Australian National University

Revised: March 1985

Abstract

In this paper we examine some existing procedures for testing linear and log-linear regression models, especially with respect to the truncation of the disturbance term in the linear model. A discrimination criterion based on two Lagrange multiplier test statistics is suggested. Some new computationally straightforward exact tests which treat the linear and log-linear regressions explicitly as non-nested models are derived. These tests are also generalized to take account of truncation in the errors of the linear model.

*The authors wish to thank Les Godfrey, J. D. Sargan and Hal White for helpful comments and discussions. An earlier version of this paper was presented to the Econometric Study Group Conference at the University of Warwick, and at the Universities of Amsterdam and Maryland. The financial support of the Research Board and the Bureau of Economic and Business Research of the University of Illinois is gratefully acknowledged.

1. Introduction

Economic theory provides information regarding the range of variables that enter into functional relationships, but it is especially weak with regard to the specific nature of those relations. It is therefore heartening to see the recent attention that is being paid to the testing of various model specifications that may arise in practice. Consider, for example, the class of data transformations involved in testing linear and log-linear regression models. Two different directions have been taken with regard to testing such models against each other. Aneuryn-Evans and Deaton (1980) treat linear and log-linear specifications as inherently non-nested and base their testing procedures on the modified likelihood ratio (LR) principle of Cox (1961, 1962). This approach is an adaptation of the classical LR test, requiring estimation under both the null and alternative models, and leads to test statistics which are very complicated to compute. A different approach has been taken by Godfrey and Wickens (1981), who base their tests on the data transformation suggested by Box and Cox (1964). Although both linear and log-linear variants may be tested by the LR method against the artificial compound model in which they are both nested, Godfrey and Wickens consider two approaches which effectively require estimation only under the null hypothesis. The first method is based on the Lagrange multiplier (LM) principle and leads to test statistics which are asymptotically chi-squared with one degree of freedom under the respective null hypotheses. A second approach appeals to the work of Andrews (1971), whose Taylor-expansion of the transformed variables leads to an exact test of the (linear) null model.

This paper has several aims. In Section 2, we consider the effects of truncation on the test procedures suggested by Godfrey and Wickens. It turns out that the LM statistic for testing the log-linear model is unaffected by truncation, whereas the LM test of the linear model is affected, in different ways, by both symmetric and unsymmetric truncation. It is shown in Section 3 that when neither the linear nor log-linear model is rejected, a choice between them may be based simply upon the difference between the two calculated LM statistics. In Section 4 we provide some new procedures for testing models in which the dependent variable is subjected to different (known) data transformations. Since these tests are based on artificial regressions which use some prior information on the specific form of the alternative model, their power properties would be expected to differ from those of the LM tests already mentioned. When the null model is log-linear, or is linear with negigible truncation, these new tests have the property of a known null distribution. In addition, the tests are conceptually straightforward and are easy to calculate. In Section 5, the tests developed are modified to incorporate truncation in the disturbances of the linear model.

2. Effects of Truncation on Some Existing Test Procedures

Consider the following two non-nested specifications

$$H_o:\quad y_t^* = f_t(x_t;\beta) + \varepsilon_{ot} \tag{1}$$

$$H_1:\quad y_t = g_t(z_t;\gamma) + \varepsilon_{1t} \tag{2}$$

in which y_t^* is a one-to-one known transformation of y_t, f_t and g_t are known (possibly) non-linear functions with arguments $(x_t;\beta)$ and $(z_t;\gamma)$,

respectively, in which the vectors of explanatory variables x_t and z_t are exogenous and the vectors of unknown coefficients β and γ are identifiable under H_o and H_1, respectively. Following Aneuryn-Evans and Deaton (1980, p. 276), we assume that

$$\varepsilon_{ot} \sim NID(0,\sigma_o^2).$$

Different transformations on y_t to obtain y_t^* impose some restrictions on the possible range of values of y_t. For example, if $y_t^* = \log y_t$, y_t must be restricted to be positive so that ε_{1t} cannot be $NID(0,\sigma_1^2)$. Aneuryn-Evans and Deaton (1980, p. 276-277) consider symmetric truncation for ε_{1t} when $y_t^* = \log y_t$, in which case the density function for ε_{1t} can be written as

$$\begin{aligned}\pi(\varepsilon_{1t}) &= \alpha(m)p(\varepsilon_{1t};\sigma_1^2), && |\varepsilon_{1t}| \leqq m\sigma_1 \\ &= 0 &&, \; |\varepsilon_{1t}| > m\sigma_1\end{aligned} \tag{4}$$

where $p(\varepsilon_t;\sigma^2)$ is the probability density function (p.d.f.) of an $N(0,\sigma^2)$ variate and $\alpha(m)$ is given by

$$\alpha(m) = \{\int_{-m}^{m} (2\pi)^{-1/2} e^{-1/2\xi^2} d\xi\}^{-1}. \tag{5}$$

For "large" values of m, $\alpha(m)$ in (5) will be close to unity and $\pi(\varepsilon_{1t})$ in (4) will be approximately $N(0,\sigma_1^2)$.

The approach adopted by Godfrey and Wickens (1981) is to nest two special forms of (1) and (2), namely

$$H_o: \; \log y_t = \beta_o + \sum_{i=1}^{k} \beta_i \log x_{it} + \varepsilon_{ot} \tag{6}$$

$$H_1: \qquad y_t = \gamma_o + \sum_{i=1}^{k} \gamma_i x_{it} + \varepsilon_{1t} \tag{7}$$

within the Box-Cox framework

$$y_t(\lambda) = \beta_o + \sum_{i=1}^{k} \beta_i x_{it}(\lambda) + \varepsilon_t \tag{8}$$

in which

$$\begin{aligned} v_t(\lambda) &= (v_t^{\lambda} - 1)/\lambda, \qquad \lambda \neq 0 \\ &= \log v_t \quad , \qquad \lambda = 0 \end{aligned}$$

and $\beta_i = \gamma_i (i > 0)$ in (8) when $\lambda = 1$. For notational convenience only, no distinction is made between the coefficients in (6) and in (8). It will be assumed that $\lambda \quad [0,\infty)$ for simplicity, implying that the parameter value lies at the boundary of the parameter space under H_o: $\lambda = 0$. This does not, however, affect the validity of the LM tests of H_o [see Moran (1971) and Chant (1974)]. If the effects of truncating ε_{1t} are small, the disturbance in (8) will be approximately $NID(0,\sigma^2)$. The LM tests of Godfrey and Wickens are obtained by regressing the unit element on the partial derivatives of the log-likelihood function of (8) with respect to (β,σ^2,λ) evaluated at the restricted estimates under the respective null models H_o: $\lambda = 0$ and H_1: $\lambda = 1$. The LM statistics are simply the sample size times the respective coefficients of multiple determination from the artificial regression with the unit element as the "dependent variable" and the partial derivatives as the "independent variables."

It is claimed by Godfrey and Wickens (1981, p. 491) that the LM statistic is unaffected by a particular form of <u>symmetric</u> truncation

of ε_{1t} in (7). Their observation was based on the fact that using the truncated p.d.f. $\pi(\varepsilon_{1t})$ in (4) simply adds the known constant $\log \alpha(m)$ to the log-likelihood function for one observation. The log-likelihood function with truncated p.d.f. is

$$\ell(\beta,\sigma^2,\lambda) = \sum_{t=1}^{T} \ell_t(\beta,\sigma^2,\lambda)$$

where

$$\ell_t(\beta,\sigma^2,\lambda) = -1/2 \log 2\pi - 1/2 \log \sigma^2 + (\lambda-1)\log y_t$$
$$- \frac{1}{2\sigma^2}\{y_t(\lambda) - \beta_o - \sum_{i=1}^{k} \beta_i x_{it}(\lambda)\}^2 + \log \alpha(m,\lambda). \quad (9)$$

If we consider a more general form of symmetric truncation within the context of the Box-Cox transformation, $\alpha(m)$ in (4) and (5) would be modified to incorporate the parameter λ. The appropriate modification of $\alpha(m)$ is given by

$$\alpha(m,\lambda) = \{\int_{-m/\lambda}^{m/\lambda} \phi(\xi)d\xi\}^{-1} \quad (10)$$

where

$$\phi(\xi) = (2\pi)^{-1/2} e^{-1/2\xi^2}.$$

Notice that (10) contains distributions corresponding to H_o: $\lambda = 0$ and H_1: $\lambda = 1$ as special cases, and that $\alpha(m,\lambda)$ will be close to unity, irrespective of the finite non-zero value for m, as $\lambda \to 0^+$. In particular, note that $\alpha(m,\lambda)$ is <u>not</u> close to unity when $\lambda = 1$ and m is not "large," since $\alpha(m,1)$ in (10) is identical to $\alpha(m)$ in (5). However, the use of (5) does not include the distributions of both the linear and

log-linear models as special cases, and hence does not exploit fully the presence of the parameter λ in the Box-Cox framework. Within this more general context, it is useful to examine whether the LM test of the linear model is affected by symmetric truncation.

The effects of truncation can be evaluated from

$$\frac{\partial \log \alpha}{\partial \lambda} = -\alpha \cdot \frac{\partial}{\partial \lambda}\left[\int_{-m/\lambda}^{m/\lambda} \phi(\xi)d\xi\right]$$

$$= -\alpha \cdot \frac{\partial}{\partial \lambda}\left[\int_{-\infty}^{m/\lambda} \phi(\xi)d\xi - \int_{-\infty}^{-m/\lambda} \phi(\xi)d\xi\right]$$

$$= -\alpha \cdot [-(m/\lambda^2)\phi(m/\lambda) - (m/\lambda^2)\phi(-m/\lambda)]$$

$$= \frac{2\,\alpha m}{\lambda^2} \cdot \phi(m/\lambda)$$

in which $\alpha \equiv \alpha(m,\lambda)$. It is easily seen that

$$\lim_{\lambda \to 1} \frac{\partial \log \alpha}{\partial \lambda} = \frac{2m\,\phi(m)}{\Phi(m)-\Phi(-m)} \tag{11}$$

and

$$\lim_{\lambda \to 0^+} \frac{\partial \log \alpha}{\partial \lambda} = 0$$

in which $\Phi(m) = \int_{-\infty}^{m} \phi(\xi)d\xi$. This shows that the LM test statistic for the log-linear model (6) will not be affected by symmetric truncation. However, for the linear model the "regressor" $\partial \ell_t/\partial \lambda$ is not the same as that given by Godfrey and Wickens because we must add the constant term $2m\,\phi(m)/(\Phi(m)-\Phi(-m))$ to it. Since the artificial regression equations

used to calculate the LM test statistics do not include an intercept, the coefficient of multiple determination with the modification given above will be different. Therefore the LM statistic for testing the linear model will be affected by symmetric truncation.

An alternative form of truncation mentioned, but not analyzed, by Aneuryn-Evans and Deaton (1980) is that of Amemiya (1973). In this case, $\alpha(m,\lambda)$ in (10) is rewritten as

$$\alpha(\beta,\sigma,\lambda) = \{\int_{-x_t'\beta/\lambda\sigma}^{\infty} \phi(\xi)d\xi\}^{-1}$$

in which $x_t'\beta = \beta_o + \sum_{i=1}^{k} \beta_i x_{it}$. It is straightforward to show that

$$\frac{\partial \log \alpha}{\partial \lambda} = \alpha \cdot (x_t'\beta/\lambda^2\sigma) \cdot \phi(-x'\beta/\lambda\sigma)$$

$$\frac{\partial \log \alpha}{\partial \beta} = \alpha \cdot (-x_t/\lambda\sigma) \cdot \phi(-x_t'\beta/\lambda\sigma)$$

$$\frac{\partial \log \alpha}{\partial \sigma} = \alpha \cdot (x_t'\beta/\lambda\sigma^2) \cdot \phi(-x_t'\beta/\lambda\sigma).$$

The above derivatives tend to zero as $\lambda \to 0^+$, whence the LM statistics for the log-linear model is unaffected. When $\lambda = 1$, however, the partial derivatives are all non-zero quantities so that the regressors required to calculate the LM statistic for the linear model, namely $\partial \ell_t/\partial\lambda$, $\partial \ell_t/\partial\beta$ and $\partial \ell_t/\partial\sigma$, will be different for the truncated model. Therefore the LM statistics for the linear and truncated linear models will be different from each other. However, when $x_t'\beta$ is large compared to σ, all these derivatives will be negligible and hence the effects of truncation will be small.

3. A Discrimination Criterion Based on LM Statistics

An especially simple criterion for discriminating between H_o and H_1 is the Sargan (1964) criterion which is based on the difference between maximized log-likelihoods when truncation of the linear model is negligible. If $\hat{\ell}_1-\hat{\ell}_o$ is the log of the likelihood ratio, H_1 is chosen if the difference is positive and H_o if it is negative. An alternative approach may be based upon the calculated values of the LM statistic under both H_o and H_1. Denote by LM_o and LM_1 the LM statistics for H_o and H_1, respectively. Let $\theta = (\beta,\sigma^2,\lambda)' \quad \Theta^{k+3}$. The asymptotic equivalence between the LM and LR statistics under the tested hypotheses and for local alternatives permits us to write

$$LM_o \overset{a}{=} LR_o = 2[\sup_{\Theta} \ell(\theta) - \sup_{\Theta|\lambda=0} \ell(\theta)]$$

and

$$LM_1 \overset{a}{=} LR_1 = 2[\sup_{\Theta} \ell(\theta) - \sup_{\Theta|\lambda=1} \ell(\theta)]$$

where $\overset{a}{=}$ denotes asymptotic equivalence. Thus, in large samples, the difference between the LM statistics is

$$LM_1 - LM_o \overset{a}{=} 2[\sup_{\Theta|\lambda=0} \ell(\theta) - \sup_{\Theta|\lambda=1} \ell(\theta)]. \qquad (12)$$

The model selection criterion implicit in (12) is that when LM_1 and LM_o are both insignificant, $H_1(H_o)$ should be chosen if LM_1-LM_o is negative (positive). Of course, a similar rule can be applied even if both LM_1 and LM_o are significant, although the selection rule will not be consistent with the outcome of the statistical test. Our suggestion of choosing the model with the <u>smaller</u> calculated LM statistic is obviously

very similar to the Sargan criterion of selecting the model with the larger maximized likelihood. Since both LM_o and LM_1 are calculated in testing H_o and H_1, respectively, the discrimination criterion based on the LM statistics is very straightforward to apply.

4. Some New Exact Non-nested Tests

The LM statistics discussed above are essentially testing the special cases $\lambda = 0$ and $\lambda = 1$ against the Box-Cox transformation given in (8). Suppose instead that a test of the (null) hypothesis H_o: $\lambda = 0$ is required to have high power against the specific non-nested alternative H_1: $\lambda = 1$, and vice-versa. In this section we develop some new and computationally straightforward tests of the linear and log-linear models against each other when truncation of the disturbance term in the linear model is negligible.

When the mean of y_t is several standard deviations above zero (i.e., when m is large for the symmetric truncation case), ε_{1t} is approximately normally distributed. For sufficiently large m, this approximation will be very close so that (6) and (7) may be supplemented with NID errors as

$$H_o: \quad \log y_t = \beta_o + \sum_i \beta_i \log x_{it} + \varepsilon_{ot}, \quad \varepsilon_{ot} \sim NID(0,\sigma_o^2) \qquad (6')$$

$$H_1: \quad y_t = \gamma_o + \sum_i \gamma_i x_{it} + \varepsilon_{1t}, \quad \varepsilon_{1t} \sim NID(0,\sigma_1^2). \qquad (7')$$

Combining the disturbances from H_o and H_1 yields

$$(1-\alpha)(\log y_t - \beta_o - \sum_i \beta_i \log x_{it}) + \alpha(y_t - \gamma_o - \sum_i \gamma_i x_{it}) = \varepsilon_t \qquad (13)$$

in which ε_t is normally, independently and identically distributed under both H_o: $\alpha = 0$ and H_1: $\alpha = 1$. Rearrangement of (13) leads to the following two artificially constructed models

$$\log y_t = \beta_o + \sum_i \beta_i \log x_{it} + \theta_o \varepsilon_{1t} + \varepsilon_t \tag{14}$$

$$y_t = \gamma_o + \sum_i \gamma_i x_{it} + \theta_1 \varepsilon_{ot} + \varepsilon_t \tag{15}$$

in which $\theta_o = -\alpha/(1-\alpha)$ and $\theta_1 = -(1-\alpha)/\alpha$. A test of H_o: $\theta_o = 0$ in (14) is equivalent to a test of $\alpha = 0$, and testing H_1: $\theta_1 = 0$ in (15) is equivalent to testing $\alpha = 1$. However, as (14) and (15) stand, H_o and H_1 are not testable because ε_{1t} and ε_{ot} are not observed variables. The approach we will adopt is to replace the disturbances from H_1 and H_o in (14) and (15), respectively, with certain estimated residuals. As will be seen, this approach has the dual advantage of testability and exactness.

In order to make the hypotheses contained in (14) and (15) testable, let us denote by $\hat{\beta}_j$ and $\tilde{\gamma}_j$ $(j = 0,1,\ldots,k)$ the OLS estimates of β_j and γ_j from (6') and (7'), respectively, so that $\widehat{\log} y_t = \hat{\beta}_o + \sum_i \hat{\beta}_i \log x_{it}$ and $\tilde{y}_t = \tilde{\gamma}_o + \sum_i \tilde{\gamma}_i x_{it}$. Now consider the artificial regressions

$$\exp(\widehat{\log}\, y_t) = \gamma_o + \sum_i \gamma_i x_{it} + \eta_{1t} \tag{16}$$

$$\log \tilde{y}_t = \beta_o + \sum_i \beta_i \log x_{it} + \eta_{ot} \tag{17}$$

in which the dependent variables in (7') and (6') have been replaced by transformations of predicted values from (6') and (7'). Denote the residuals from (16) and (17) by

$$\hat{\eta}_{1t} = \exp(\widehat{\log y_t}) - \hat{\gamma}_o - \sum_i \hat{\gamma}_i x_{it} \tag{18}$$

$$\tilde{\eta}_{ot} = \log \tilde{y}_t - \tilde{\beta}_o - \sum_i \tilde{\beta}_i \log x_{it} \tag{19}$$

and substitute (18) and (19) into (14) and (15), respectively, in place of the unobserved ε_{1t} and ε_{ot}. The tests of H_o and H_1 are the t-ratios for the OLS estimates of θ_o and θ_1 in

$$\log y_t = \beta_o + \sum_i \beta_i \log x_{it} + \theta_o \hat{\eta}_{1t} + \varepsilon_t \tag{20}$$

$$y_t = \gamma_o + \sum_i \gamma_i x_{it} \qquad + \theta_1 \tilde{\eta}_{ot} + \varepsilon_t . \tag{21}$$

Since $\hat{\eta}_{1t}$ is a function of $\widehat{\log y_t}$ through (16), $\hat{\eta}_{1t}$ is independent of the OLS residuals from (20) under H_o; similarly, $\tilde{\eta}_{ot}$ is independent under H_1 of the OLS residuals from (21) because it is a function of $\tilde{y}_t$ through (17). Therefore, the results of Milliken and Graybill (1970) may be used to show that the relevant t-ratios for $\hat{\theta}_o$ and $\tilde{\theta}_1$ from (20) and (21), respectively, both have the exact t-distribution with (T-k-2) degrees of freedom. Strictly speaking, of course, the test of the linear model in (21) has only the <u>approximate</u> t-distribution under H_1 because ε_{1t} is only approximately normally distributed for large m.

The test procedures are quite straightforward to implement. For example, in order to test H_1 in (7'), we require only the following two simple steps:

(i) Replace log y_t in (6') with log $\tilde{y}_t$ (as in (17)), and obtain the residuals $\tilde{\eta}_{ot}$ (as in (19));

(ii) Test the significance of $\tilde{\eta}_{ot}$ when it is included in H_1 (as in (21)), and reject (do not reject) H_1 if $\tilde{\theta}_1$ is (is not) significantly different from zero.

It is worth noting that the nesting procedure used in (13) is arbitrary. We could, of course, have considered a weighting scheme of the form

$$(1-\alpha)\exp(\log y_t - \beta_o - \sum_i \beta_i \log x_{it}) + \alpha(y_t - \gamma_o - \sum_i \gamma_i x_{it}) = \varepsilon_t \qquad (22)$$

so that the test of H_1 would be the t-ratio of $\tilde{\theta}_1$ in

$$y_t = \gamma_o + \sum_i \gamma_i x_{it} + \theta_1 \exp(\tilde{\eta}_{ot}) + \varepsilon_t \qquad (23)$$

instead of in (21). Although the t-ratios for $\tilde{\theta}_1$ from (21) and (23) both have the exact t(T-k-2) distribution under H_1, and hence are asymptotically equivalent under the null and against local alternatives, the power of H_1 may well differ between (21) and (23). Therefore, it would be useful to compute the test statistics from both equations for purposes of comparison. Notice that we cannot use the weighting scheme

$$(1-\alpha)(\log y_t - \beta_o - \sum_i \beta_i \log x_{it}) + \alpha \log(y_t - \gamma_o - \sum_i \gamma_i x_{it}) = \varepsilon_t \qquad (24)$$

instead of (13) and (22), and the testing equation

$$\log y_t = \beta_o + \sum_i \beta_i \log x_{it} + \theta_o \log(\hat{\eta}_{1t}) + \varepsilon_t \qquad (25)$$

rather than (20), because there is a non-zero probability that $(y_t - \gamma_o - \sum_i \gamma_i x_{it})$ in (24) and $\hat{\eta}_{1t}$ in (25) may be negative for some t.

A related problem arises in the derivation of the Andrews test by Godfrey and Wickens. These tests of H_o and H_1 are based on artificial regressions similar to (20) and (21), in which $\hat{\eta}_{1t}$ and $\tilde{\eta}_{ot}$ are replaced by

$$\hat{q}_{1t} = \tfrac{1}{2}[(\log \hat{y}_t)^2 - \sum_i \hat{\beta}_i(\log x_{it})^2]$$

and

$$\tilde{q}_{ot} = [(\tilde{y}_t \log \tilde{y}_t - \tilde{y}_t + 1) - \sum_i \tilde{\gamma}_i(x_{it} \log x_{it} - x_{it} + 1)]$$

respectively. The appropriate tests of H_o and H_1 are simply the t-ratios of the estimated coefficients of $\hat{q}_{1t}$ and $\tilde{q}_{ot}$, respectively, with the test of $H_o(H_1)$ being distributed exactly (approximately exactly) as $t(T-k-2)$ under the respective null hypotheses. Thus, both the Andrews test and the new test derived in this section have the same size in small samples and are, of course, asymptotically equivalent under the tested model. However, their respective powers may differ in finite samples when the alternative is "true."

It is useful to mention three aspects of the tests developed above. First, it has been assumed that truncation of the linear moel is negligible, so that ε_{1t} is normal. There is, therefore, a non-zero probability of obtaining negative values of y_t (and hence of $\tilde{y}_t$) for <u>some</u> t, in which case $\tilde{\eta}_{ot}$ may not be obtainable. In such circumstances, an exact test of the linear model may be unavailable. Similar views were expressed by Godfrey and Wickens (1981, p. 493) in relation to the Andrews test of the linear model. Second, the predicted values $\exp(\log \hat{y}_t)$ and $\log \tilde{y}_t$ may be used to replace $\hat{\eta}_{1t}$ and $\tilde{\eta}_{ot}$, respectively, in (20) and (21). The exactness of the tests of H_o and H_1 is unaffected, although the small sample power characteristics may differ when predictions under the null replace the calculated residuals. Third, while the exactness of the tests depends upon the normality assumption, their asymptotic validity does not.

5. Tests in the Presence of Truncation

When m is not large so that truncation of the linear model may not reasonably be ignored, it is convenient to assume that ε_{1t} is distributed as truncated normal in the range $(-m_t,\infty)$, where $m_t = \gamma_o + \sum_i \gamma_i x_{it}$ (see Amemiya, 1973). The p.d.f. of ε_{1t} can then be written as

$$\pi(\varepsilon_{1t}) = [F_t(2\pi)^{1/2}\sigma_1]^{-1} e^{-1/2\varepsilon_{1t}^2/\sigma_1^2}$$

where

$$F_t = \int_{-\infty}^{m_t} [(2\pi)^{1/2}\sigma_1]^{-1} e^{-1/2\xi^2/\sigma_1^2} d\xi .$$

It is easily seen that $E(\varepsilon_{1t}) = \sigma_1^2(f_t/F_t)$, where

$$f_t = [(2\pi)^{1/2}\sigma_1]^{-1} e^{-1/2m_t^2/\sigma_1^2}.$$

Now consider the augmented version of H_1 (see Olsen, 1980, p. 1817) given as

$$H_1^*: \quad y_t = \gamma_o + \sum_i \gamma_i x_{it} + \delta(f_t/F_t) + \varepsilon_{1t}^* \tag{26}$$

where $\delta = \sigma_1^2$ and $\varepsilon_{1t}^* = \varepsilon_{1t} - \delta(f_t/F_t)$. Note that the disturbance term in (26) is heteroscedastic with zero mean. Therefore, for purposes of testing, H_1 can be replaced by H_1^*. The quantity (f_t/F_t) is, however, unknown because it depends on the unknown parameters $(\gamma_o,\gamma_i,\sigma_1^2)$. These parameters may be replaced by their maximum likelihood estimates $(\tilde{\gamma}_o,\tilde{\gamma}_i,\tilde{\sigma}_1^2)$, from which we can obtain $(\tilde{f}_t/\tilde{F}_t)$. A suitably modified form of H_1 is therefore given by

$$H_1^{**}: \quad y_t = \gamma_o + \sum_i \gamma_i x_{it} + \delta(\tilde{f}_t/\tilde{F}_t) + \varepsilon_{1t}^{**}. \tag{27}$$

Equation (27) is equivalent to (7'), except for the appearance of the additional regressor $(\tilde{f}_t/\tilde{F}_t)$ and the fact that the disturbance term is now heteroscedastic and serially correlated since $\varepsilon^{**}_{1t} = \varepsilon^{*}_{1t} + \delta(f_t/F_t) - \delta(\tilde{f}_t/\tilde{F}_t)$. This will lead to a modification of our test procedure. For testing the log-linear model, we would replace equation (16) with

$$\exp(\widehat{\log y_t}) = \gamma_o + \sum_i \gamma_i x_{it} + \delta(\tilde{f}_t/\tilde{F}_t) + \eta_{1t}$$

and the analysis would proceed as before. The t-ratio for $\hat{\theta}_o$ in the equivalent of (20) is asymptotically distributed under H_o as N(0,1). In order to test the linear model, the <u>truncated</u> linear model is first estimated to obtain the predicted values $\tilde{y}_t$, which are then used in equation (17) to obtain $\tilde{\eta}_{ot}$ as in (19). Finally, (21) is replaced by

$$y_t = \gamma_o + \sum_i \gamma_i x_{it} + \delta(\tilde{f}_t/\tilde{F}_t) + \theta_1 \tilde{\eta}_{ot} + \varepsilon_t. \qquad (28)$$

In this case it should be noted that the test is not exact because the tested model is not actually H_1, but rather H_1^{**} (i.e., H_1 modified to account for truncation). Since the disturbance in (28) is both heteroscedastic and serially correlated, for valid inferences the consistent covariance matrix estimator of White and Domowitz (1984) should be used for testing the hypothesis H_1^{**}: $\theta_1 = 0$. The appropriate t-ratio for $\tilde{\theta}_1$ in (28) will be asymptotically distributed under H_1 as N(0,1).

It would be interesting to investigate how these modifications to incorporate truncation in the linear model, as well as the LM statistics discussed in Section 2, will affect inferences. It is worth observing that in many empirical studies the log-linear model has been found to

be superior to the linear model when truncation has been (rightly or wrongly) ignored. The results might well be different, at least for models based on survey data, if a suitably truncated linear model were to be tested against a log-linear model.

6. Summary

This paper addressed several issues associated with testing linear and log-linear regression models against each other. The analysis in Section 2 focused attention on symmetric and unsymmetric truncation of the errors in the linear model, and showed how the Lagrange multiplier test of the linear model is affected when a general form of truncation is used. The effect of the truncation depends on the relative magnitudes of the regression component and the error variance of the linear model. In Section 3 it was shown that the Lagrange multiplier statistics used to test the hypotheses can also be used to discriminate between them. Some new computationally straightforward exact tests which treat the linear and log-linear regressions explicitly as non-nested models were derived in Section 4. These tests may easily be generalized to handle different data transformations of the dependent variable. While the exactness of the tests depends upon the normality assumption, their asymptotic validity does not. These tests were modified in Section 5 to incorporate truncation in the errors of the linear model.

References

Amemiya, T. (1973), "Regression Analysis when the Dependent Variable is Truncated Normal," Econometrica 41, 997-1016.

Andrews, D. F. (1971), "A Note on the Selection of Data Transformations," Biometrika 58, 249-254.

Aneuryn-Evans, G. and A. S. Deaton (1980), "Testing Linear Versus Log-arithmic Regression Models," Review of Economic Studies 47, 275-291.

Box, G. E. P. and D. R. Cox (1964), "An Analysis of Transformations," Journal of the Royal Statistical Society B 26, 211-252.

Chant, D. (1974), "On Asymptotic Tests of Composite Hypotheses in Nonstandard Conditions," Biometrika 61, 291-298.

Cox, D. R. (1961), "Tests of Separate Families of Hypotheses," Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability 1 (Berkeley, University of California Press), 105-123.

Cox, D. R. (1962), "Further Results on Tests of Separate Families of Hypotheses," Journal of the Royal Statistical Society B 24, 406-424.

Godfrey, L. G. and M. R. Wickens (1981), "Testing Linear and Log-linear Regressions for Functional Form," Review of Economic Studies 48, 487-496.

Milliken, G. A. and F. A. Graybill (1970), "Extensions of the General Linear Hypothesis Model," Journal of the American Statistical Association 65, 797-807.

Moran, P. A. P. (1971), "Maximum-Likelihood Estimation in Non-standard Conditions," Proceedings of the Cambridge Philosophical Society 70, 441-450.

Olsen, R. (1980), "A Least Squares Correction for Selectivity Bias," Econometrica 48, 1815-1820.

Sargan, J. D. (1964), "Wages and Prices in the United Kingdom: A Study in Econometric Methodology," in P. E. Hart, G. Mills and J. K. Whitaker (eds), Econometric Analysis for National Economic Planning (London, Butterworths), 25-63.

White, H. and I. Domowitz (1984), "Nonlinear Regression with Dependent Observations," Econometrica 52, 143-161.